Acces to Grey Content : An Analysis of Grey Literature based on Citation and Survey Data : A Follow-up Study

Dominic J. Farace, Jerry Frantzen, Joachim Schöpfel, Christiane Stock, Albert Boekhorst

Access to Grey Content: An Analysis of Grey Literature based on Citation and Survey Data: A Follow-up Study

Dominic J. Farace
GreyNet, Netherlands

Jerry Frantzen
Boer & Croon, Netherlands

Joachim Schöpfel and Christiane Stock
INIST-CNRS, France

Albert K. Boekhorst
UvA, Netherlands; UP, South Afirca

Introduction

Grey literature, an area of interest to special librarians and information professionals, can be traced back a half-century. However, grey literature as a specialized field in information studies is less than a decade old. At GL'97 in Luxembourg, grey literature was redefined "as information produced on all levels of government, academics, business and industry in electronic and print formats not controlled by commercial publishers (i.e. where publishing is not the primary activity of the producing body)." The subject area was broadened and the need for continuing research and instruction pursued. The results of an online survey carried out in 2004 compared with survey results a decade prior indicate two changes: (1) a move to more specialization in the field of grey literature and (2) a move to more balance in activities related to research and teaching as compared with the processing and distribution of grey literature. It is not that the activities of processing and distribution are today of less concern, but technological advances and the Internet may have made them less labour intensive. The burden that grey literature poised to human resources and budgets appears to have been reduced enough that the benefits of its content is discovered. And this discovery of a wealth of knowledge and information is the onset to further research and instruction in the field.

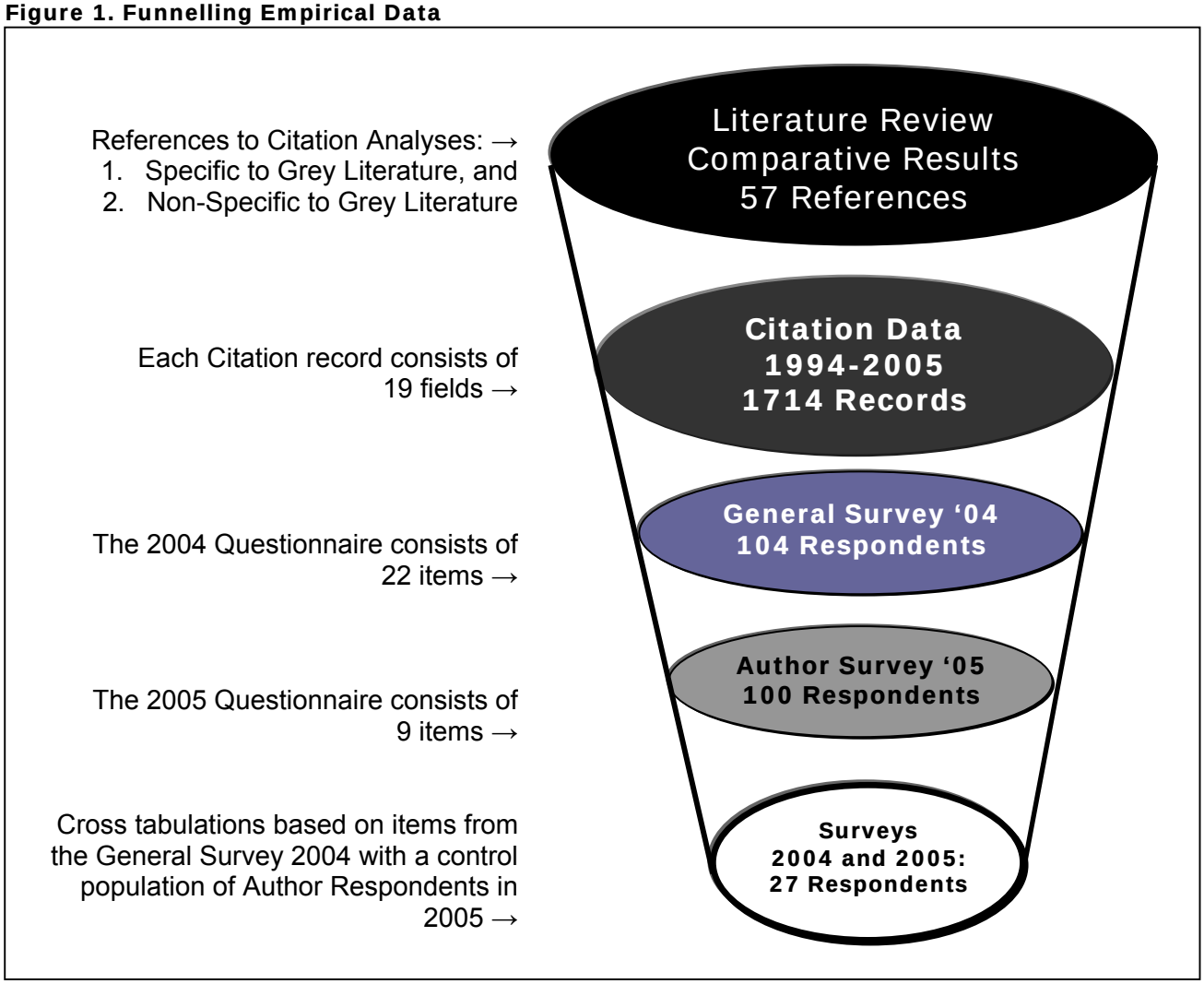
Research Goal

The idea behind this study is that - by using the same pool of authors - survey data linked to citation data will allow for a clearer demonstration of the impact of their research, where only part of the impact is covered by citation analysis alone. Hopefully, the new combined results will provide a better profile of these meta-authors, who are also the source of GreyNet's knowledge and information base. This could lead to the subsequent development of information policies and services that are more in line with the needs of authors and researchers, whereby their results would become even more accessible well beyond the grey circuit.

Research Plan

This research is a follow-up to two projects carried out in 2004. One was a citation analysis based on the published papers in the GL Conference Proceedings and the other was a general survey, which dealt with the response of information professionals to key issues and topics in the field of grey literature. In this study, we seek not only to update and integrate the data from the citation analysis but at the same time to introduce the instrument of an author survey in order to better assess the work and expectations of those who are actually doing research and authoring papers on the topic of grey literature. These are referred to as the meta-authors of grey literature.

Figure 1. Funnelling Empirical Data



I. Results of Citation Analysis 1994-2005

It is not the intention here to repeat the findings of last year's research [22], which was a more textual account and analysis, but rather suffice to provide another format in order to present the cumulative results of the research. To this end, tables and their explanations are emphasized. Further, we find that once the citation database had been updated with the records from the GL6 Conference Proceedings (2005), new trends and developments can be identified. And, it is these that could have a marked influence on access to grey content issuing from the conference series.

Table 1.
General Citation Data

Conf No.	No. of papers	Papers without citations	No. of papers with citations	Total No. of citations	Average No. of citations per paper
1	37	9	28	345	12.3
2	25	4	21	247	11.8
3	29	10	19	275	14.5
4	28	4	24	250	10.4
5	20	3	17	227	13.4
6	24	0	24	370	15.4
Total:	**163**	**30**	**133**	**1714**	**13.0**

Unlike previous conferences in the series, all of the conference papers in the GL6 Proceedings without exception contained references. These same proceedings claim the highest number of citations (370) irrespective of the number of conference papers. And, these proceedings maintain the highest average number of citations (15.4) per conference paper.

Table 2.
Standard or Hyperlink Citations

Conf No.	Citations total	Standard citations	Hyperlinked citations	Explanative notes	Name & address
1	345	322	0	15	8
2	247	243	2	2	0
3	275	207	63	4	1
4	250	160	76	14	0
5	227	155	67	5	0
6	370	195	162	12	1
Total:	**1714**	**1282**	**370**	**52**	**10**

The GL6 Conference Proceedings not only had the highest average number of hyperlinked citations compared with standard citations from previous conferences in the series, but also the total number nearly doubled compared with that of the year prior. However, this increase in hyperlinked citations was not at the cost of standard citations, because the overall number of citations per conference paper had increased. Noticeably, what did not increase was the quality of the hyperlinked citations. A considerable number of which only show a URL without further description of the source. This may bear out what Chu[20] infers by hyperlinked citations being different from standard citations in that they point more to resources rather than they support or refute academic research.
Further what we find in the search of the citation database, but which is not shown in the table above, is the increase of citations to grey publications compared to commercial publications. This distinction is based on document type and can again be explained by the increase in hyperlinked citations, which referred mostly to WebPages [1] and Web papers available through non-commercial publishers. It is then important to researchers and librarians than an accurate link between publications and their references are made. [2]

Table 3.
Serial Citations

Conf No.	No. of papers with citations	Citations total	No. of serial citations	Maximum Citations per paper	Minimum Citations per paper
1	28	345	0	69	1
2	21	247	23	73	1
3	19	275	17	62	1
4	24	250	13	27	1
5	17	227	26	31	2
6	24	370	58	35	1
Total :	**133**	**1714**	**137**		

Serial citations (i.e. citations to previous conference papers in the GL Series) have not only doubled in total number every year for the past three conferences but also show an average increase of more than 5% for each of the same past three consecutive conferences - from 5% in the 4th to 11.5% in the 5th to 15.7% in the 6th. This may indicate not only more access to previous conference papers and/or proceedings in the GL-Series but also further use and application of research results originating from within this Conference Series.

Table 4.
Self-Citations

Conf no	No. of papers	Citations total	No. of Self-citations	No. of Non-Self citations	Not applicable*
1	28	345	42	231	72
2	21	247	15	189	43
3	19	275	19	175	81
4	24	250	18	134	98
5	17	227	40	128	59
6	24	370	37	221	112
Total:	**133**	**1714**	**171**	**1078**	**465**

The number of self-citations in the GL6 Proceedings (10%) appears to be declining to the level of the first 4 conferences, which together averaged 8%. These conferences were prior to the 2000-2003 break in the series. After its relaunch with the 5th conference, the meta-authors may have had to rely in that conference year on their own findings to substantiate arguments, since they did not have ready access to previous conference papers nor to GreyNet, the Grey Literature Network Service, which was also dormant in that same four-year period.

Table 5.
Age of Citations

Conf no	Citations total	No. of citations in Year of Conf.	No. Minus 1 Year	No. Minus 2 Years	No. Minus 3 Years	Earliest year of cited work
1	345	83	37	22	13	1949
2	247	60	71	15	26	1944
3	275	86	41	18	22	1945
4	250	89	32	26	13	1886
5	227	64	28	21	16	1949
6	370	177	28	19	11	1896
Total	**1714**	**559**	**237**	**121**	**101**	

Another significant figure from the citation data of GL6 is the sharp increase in the number of citations dated the same year of the conference, which was 47,8%. Looking at the overall average of the first 5 conferences in the GL Series, the average was 28,4%. Once again, this increase not only illustrates a trend in research to cite current work but is also influenced by the ratio of hyperlinked citations that carry the date in which the conference is held.
While it is beyond the scope of this current paper, future analysis of the citation data once the GL7 records have been entered in the database may reveal further trends and distinctions between standard and hyperlinked citations as they impact and influence work by meta-authors in the field of grey literature.

II. Results of the Author Survey 2005

If we now turn to the results of the author survey carried out in 2005, a brief word on the population of the respondents show that they are all past or present authors in the GL-Conference Series. The total population of these meta-authors (i.e. informational professionals working in the field of grey literature and doing research and authoring papers) since the start of the GL conference series in 1993 is roughly 230. Initially, there were 103 respondents to this online survey. However, three of them withdrew their content submission to GL7 and subsequently were deleted from the survey bringing the total number of respondents to an even hundred.

Table 6.
Continent where the Author lives and works

	Frequency	Percent
North America	35	35,0
Europe	52	52,0
Asia	8	8,0
Other	5	5,0
Total	100	100,0

Across the board, whether looking at citation data such as cited works and citing authors or whether looking at the respondents to the general survey in 2004 or this Author Survey in 2005, North America and Europe account for 85% to 90% of global activity in the field of grey literature.

Table 7.
Citation Style for Grey Literature would be of benefit for the author's work?

	Frequency	Percent
Yes	55	55,0
No	21	21,0
Depends	11	11,0
NA	13	13,0
Total	100	100,0

While more than 50% of the authors respond with a simple yes to this open question, another 11% would be inclined to such guidelines as long as it would not complicate and duplicate their work at hand. Such guidelines should be in general use supported by a global community and in place for multiple types of grey literature.

Table 8.
Commercial publisher accepted one or more of their works?

	Frequency	Percent
Yes	49	49,0
No	38	38,0
Depends	5	5,0
NA	8	8,0
Total	100	100,0

Nearly 50% of the authors had one or more of their manuscripts accepted by a commercial publisher. Another 5% is unclear - depending on crossover situations - where a grey publisher was taken over by a commercial publisher or where a publication has moved into the realm of OAI.

Table 9.
Author has published on other topics than Grey Literature?

	Frequency	Percent
Information Science	59	59,0
Other subjects	25	25,0
NA	16	16,0
Total	100	100,0

While 84% of the authors published on other topics than grey literature, 59% of them remained within the field of information science. The other 25% published in a variety of different fields in the natural sciences, social sciences, and humanities.

Table 10.
Author's view on Open Access

	Frequency	Percent
Positive (unqualified)	66	66,0
Positive (qualified)	29	29,0
NA	5	5,0
Total	100	100,0

Only 5% of the authors did not respond to this question, while 95% are favourable to Open Access. A near two-thirds sufficed with a simple statement, while 29% provided more lengthy and qualified arguments for their position. Their positions ranged the full gamut from views held and published by the Wellcome Trust* to those of The Royal Society**.

III. Comparative Results of Survey and Citation Data

In the first two parts of this paper, we looked separately at results of citation data and survey data. It is our intention in this final part of the paper to present some comparative results, as they appear from cross-tabulations of the 2004 and 2005 Surveys and the extent to which other combined data lend themselves to empirical observation.
A selection of 5 items from the 2004 Survey was made in an effort to determine if differences exist in the responses between those informational professionals simply working in the field of grey literature and the meta-authors (i.e. those who are both working in this field of information and who are also doing research and authoring publications on the topic of grey literature). The meta-authors in this research totalled twenty-seven. They in fact were the population of respondents, who completed both the 2004 General Survey and the 2005 Author Survey. Once the five items for cross-tabulation were chosen, it was then necessary to delete the 27 respondents from the 2004 Survey population so that the results of the one group would not influence the outcome of the other. This resulted in two groups of respondents having survey populations of 77 and 27 respectively. A check shows that the total number of respondents for each item is 104. However, due to some changes in the recoding of responses, minor discrepancies appear in the subtotals. Since these are only indicative results, we choose to share the findings as they are calculated in the tables below.

Table 11.
The average net-user should at least recognise the term 'grey literature'

	Information Professionals	
	Non-Meta Author	Meta-author
Depends	**0**	**7**
	,0%	25,9%
NA	**8**	**0**
	10,4%	,0%
No	**29**	**7**
	37,7%	25,9%
Yes	**40**	**13**
	51,9%	48,1%
Total	**77**	**27**
	100,0%	100,0%

On this open-ended item, we see almost an inverse relationship between the meta-authors and the non-meta authors regarding the percentage of no-answers (NA) and qualified statements (Depends). While there was little variation in the percentage that agreed, there was almost a 12-percentage point difference in those who disagreed. The Meta-authors are more convinced that the term grey literature belongs in the vocabulary of net-users.

* http://www.wellcome.ac.uk/print/wtd002766_print.html
** http://www.royalsoc.ac.uk/page.asp?id=3882

Table 12.
Grey Literature should be free to access

	Information Professionals	
	Non-Meta Author	Meta-author
Depends	**13**	**7**
	16,9%	25,9%
NA	**7**	**1**
	9,1%	3,7%
No	**12**	**2**
	15,6%	7,4%
Yes	**45**	**17**
	58,4%	63,0%
Total	**77**	**27**
	100,0%	100,0%

While the majority of both groups of respondents favour free access to grey literature, the meta-authors lead by 5-percentage points in their agreement on this particular survey item.

Table 13.
Grey Literature should be free of charge

	Information Professionals	
	Non-Meta Author	Meta-author
Depends	**17**	**10**
	22,1%	37,0%
NA	**6**	**2**
	7,8%	7,4%
No	**14**	**4**
	18,2%	14,8%
Yes	**40**	**11**
	51,9%	40,7%
Total	**77**	**27**
	100,0%	100,0%

Over half of the Non-meta authors feel that grey literature should be free of charge, while the meta-authors were not in the majority on this standpoint. Instead, they chose more often to qualify their response allowing for differences in the sector in which grey literature is produced, the size of the corporate author and/or producing body, financial position, etc.

Table 14.
Grey Literature itself constitutes a field in information studies

	Information Professionals	
	Non-Meta Author	Meta-author
Depends	**4**	**4**
	5,2%	14,8%
NA	**7**	**1**
	9,1%	3,7%
No	**14**	**2**
	18,2%	7,4%
Yes	**52**	**20**
	67,5%	74,1%
Total	**77**	**27**
	100,0%	100,0%

While the majority of both groups of respondents agree that grey literature constitutes a specialized field in information studies, it is not surprising to find that the meta-authors carry almost a 7-percentage point lead on this survey item.

Table 15.
The Luxembourg Convention on Grey Literature still holds

	Information Professionals	
	Non-Meta Author	Meta-author
Depends	**7**	**2**
	9,1%	7,4%
NA	**4**	**2**
	5,2%	7,4%
No	**10**	**4**
	13,0%	14,8%
Yes	**56**	**19**
	72,7%	70,4%
Total	**77**	**27**
	100,0%	100,0%

Not only did both groups of respondents overwhelming favour the current definition of grey literature known as the 'Luxembourg Convention' but this survey item also shows the least amount of variation in percentages between the two groups.

IV. Summary of Findings and Conclusion

In order to be clear on the results, which are based on different types of data applied in this study, separate subheadings are used below.

Based on Citation Data:

- Hyperlink citations are rapidly gaining ground on standard citations
- Hyperlink citations tend to increase the total number of citations in a conference paper
- Hyperlink citations are also increasing the number of references to grey literature
- Self-citations are decreasing, while serial citations are increasing
- Nine of the top-ten cited authors are also meta-authors in the GL Conference Series

Based on Survey Data:

- Nearly half of the meta-authors also make use of commercial publishers
- More than three quarters of the meta-authors also publish on other topics than GL
- Without reservation, nearly two-thirds of the meta-authors favour OAI
- However, nearly one-third of the meta-authors provide qualified statements on OAI, thus requiring GreyNet to further analyse these responses before rendering a position statement[***].

Based on Comparative Data:

Differences of opinion were uncovered between meta-authors and non-meta authors:

- Meta-authors were significantly more inclined to qualify their statements
- Non-meta authors were significantly more inclined not to respond to a given question
- Analysis of the top-5 types of grey literature resulting from both the citation data and the survey data show that four of them are the same, namely: conference papers, journal articles, reports, and WebPages.
- However, significant differences appear in the meta-authors' production and use of these types of grey literature. Reports are first to be produced, while fourth in line to be cited.

In close, the literature review uncovered a wealth of citation formats available for grey literature. However, the differences in formats and uses of hyperlinked and standard citations require further research and development. Perhaps the 'Nancy Style' proposed at GL7 will offer a framework for this and other best practices in the field of grey literature.

[***] GreyNet's position on OAI is scheduled for publication in the Editor's Note of The Grey Journal, TGJ volume 2, number 1, Spring 2006. – ISSN 1574-1796.

References and Sources

This is a follow-up study; references from the earlier study are not repeated here. They can be found in Ref. 22, below.

Bibliographic References Specific to Grey literature

1. Davis, Philip M. (2002), The effect of the Web on undergraduate citation behavior: a 2000 update. - In: Collection & Research Libraries, v. 63, n. 1, pp. 53-60.
2. Morrisey, Locke J. (2002), Bibliometric and bibliographic analysis in an era of electronic scholarly communication. - In: Science & Technology Libraries, v. 22, n. 3-4, pp. 149-160.
3. Kraus, Joseph R. (2002), Citation patterns of advanced undergraduate students in Biology, 2000-2002. - In: Science & Technology Libraries, v. 22, n. 3/4, pp. 161-179.
4. Brown, Cecelia (2003), The role of electronic preprints in chemical communication: analysis of citation, usage, and acceptance in the journal literature. - In: Journal of the American Society for Information Science and Technology, v. 54, n. 5, pp. 362-371.
5. Pelzer, Nancy L. and William H. Wiese (2003), Bibliometric study of grey literature in core veterinary medical journals. - In: Journal of the Medical Library Association, v.91, n.4, pp. 434-441.
6. Hasbrouck, La Mar, Joanna Tallano, Jon Mark Hirshon, and Andrew L. Dannenberg (2003), Use of epidemiology in clinical medical publications, 1983-1999: a citation analysis. - In: American Journal of Epidemiology, v.157, n.5, pp. 399-408.
7. Garg, K.C. (2003), An overview of cross-national, national, and institutional assessment as reflected in the international journal Scientometrics. - In: Scientometrics, v. 56, n. 2, pp. 169-199.
8. Hawkins, Donald T., Signe E. Larson, and Bari Q. Caton (2003), Information science abstracts: tracking the literature of information science. Part 2: A new taxonomy for information science /. In: Journal of the American Society for Information Science and Technology, v. 54, n. 8, pp. 771-781.
9. Sondergaard, Trine, Jack Andersen, and Birger Hjorland (2003), Documents and the communication of scientific and scholarly information. - In: Journal of Documentation, v. 59, n. 3, pp. 278-320.
10. Valderrama Zurian, J.C., R. Aleixandre, and M. Castellano (2004), Citation-count analysis in Addiction 2001. - In: Addiction, v. 99, n. 3, pp. 387-388.
11. Waugh, C. Keith and Margie Ruppel (2004), Citation analysis of dissertation, thesis, and research paper references in workforce education and development. - In: The Journal of Academic Librarianship, v. 30, n. 4, pp. 276-284.
12. Schaffer, Thomas (2004), Psychology citations revisited: behavioural research in the age of electronic resources. - In: The Journal of Academic Librarianship, v. 30, n. 5, pp. 354-360.
13. Ullah, Mohd. Furqan, S.S. Kanwar, and Pradeep Kumar (2004), A quantitative analysis of citations of research reports published by National Institute of Hydrology, Rorkee. - In: Annals of Library and Information Studies, v. 51, n. 3, pp. 108-115.
14. Sharif, Muhammad Ashraf and Khalid Mahmood (2004), How economists cite literature: citation analysis of two core Pakistani economic journals. - In: Collection Building, v. 23, n. 4, pp. 172-176.
15. Valderrama, Rafael Juan Carlos, José Maria Desantes, and Antonio J. Torregrosa (2004), Identification of information sources and citation patterns in the field of reciprocating internal combustion engines. - In: Scientometrics, v. 59, n. 3, pp. 321-336.
16. Beile, Penny M., David N. Boote, and Elizabeth K. Killingsworth, E.K. (2004), A microscope or a mirror? : A question of study validity regarding the use of dissertation citation analysis for evaluating research collections. - In: The Journal of Academic Librarianship, v. 30, n. 5, pp. 347-353.
17. Bremholm, Tony L. (2004), Challenges and opportunities for bibliometrics in the electronic environment: the case of the Proceedings of the Oklahoma Academy of Science. - In: Science & Technology Libraries, v. 25, n.1-2, pp. 87-107.
18. Chen, Chaomei and Diana Hicks (2004), Tracing knowledge diffusion / In: Scientometrics, v. 59, n. 2, pp. 199-211.
19. Swygart-Hobaugh, Amanda J. (2004), A citation analysis of the quantitative/qualitative methods debate's reflection in sociology research: implications for library collection development. - In: Library Collections, Acquisitions, & Technical Services, v. 28, n. 2, pp. 180-195.
20. Chu, Heting (2005), Taxonomy of inlinked Web entities: what does it imply for webometric research? - In: Library & Information Science Research, v. 27, n. 1, pp. 8-27.
21. Bollen, Johan, Herbert van de Sompel, Joan A. Smith, and Rick Luce (2005), Toward alternative metrics of journal impact: a comparison of download and citation data. - In: Information Processing & Management, v. 41, n. 6, pp. 1419-1440.
22. Schöpfel, J., C. Stock D.J. Farace and J. Frantzen (2005), Citation Analysis and Grey Literature: Stakeholders in the Grey Circuit. - In: GL6 Conference Proceedings : Sixth International Conference on Grey Literature : Work on Grey in Progress, 6-7 December 2004. – Amsterdam : TextRelease, 2005, p. 55-63. – ISBN 90-77484-04-3.

Bibliographic References Non-specific to Grey literature

23. Davis Herring, Susan (2002), Use of electronic resources in scholarly electronic journals: a citation analysis. - In: College & Research Libraries, v. 63, n. 4, pp. 334-340.
24. Kotiaho, Janne S. (2002), Ethical considerations in citing scientific literature and using citation analysis in evaluation of research performance. - In: Journal of Information Ethics, v. 11, n. 2, pp. 10-16.
25. Kostoff, Ronald N. (2002), Citation analysis of research performer quality. - In: Scientometrics, v. 53, n. 1, pp. 49-71.
26. Zhao, Dangzhi and Elisabeth Logan (2002), Citation analysis using scientific publications on the Web as data sources: a case study in the XML research area. - In: Scientometrics, v. 53, n. 4, pp. 449-472.
27. Cronin, Blaise and Debora Shaw (2002), Identity-creators and image-makers: using citation analysis and thick description to put authors in their place. - In: Scientometrics, v. 53, n. 1, pp. 31-49.
28. Koehler, Wallace (2002), Nightmares in citation analysis. - In: Reference and User Services Quarterly, v. 42, n. 1, pp. 41-42.

Sources used for the Collection of Citation Data

29. Farace, D.J. and J. Frantzen (1994-….), Conference proceedings on grey literature: GL-conference series. – ISSN 1386-2316
Amsterdam : GreyNet/TransAtlantic, 1994-2000; Numbers 1-4
Amsterdam : GreyNet/TextRelease, 2003-2005; Numbers 5-6

Author Information

Dominic J. Farace is Director of TextRelease, an Amsterdam based information bureau specializing in grey literature and networked information. He is a native Louisianan and holds two degrees in sociology from Creighton University (BA) and the University of New Orleans (MA). His doctoral dissertation in social sciences is from the University of Utrecht, The Netherlands, where he has lived and worked for the past twenty-nine years. After six years heading the Department of Documentary Information at the Royal Netherlands Academy of Arts and Sciences (Swidoc/KNAW), he founded GreyNet, Grey Literature Network Service, in 1993 and has since been responsible for the international GL-Conference Series. In this capacity, he serves as Program and Conference Director as well as managing editor of the conference proceedings. Since 2004, he is a Guest Lecturer on Grey Literature in the Masters Program at the University of Amsterdam, Faculty of Humanities. He is Guest Editor for PRQ, Publishing Research Quarterly and the editor of The Grey Journal. Email: info@greynet.org

Jerry Frantzen graduated in 1999 from the College of Amsterdam in Library and Information Science. He is presently employed with Boer and Croon Strategy and Management Group in Amsterdam and is technical editor of The Grey Journal

Joachim Schöpfel graduated from the University of Hamburg in 1984. A research assistant and lecturer at the University of Hamburg, Department of Developmental and Educational Psychology, from 1985 to 1990, he obtained his Ph.D. from the same university in 1992. He is presently head of the library and document delivery department at the French Institute of Scientific and Technical Information and teaches Culture and Society (1992-2001) and Documentation (from 2001 on) at the University of Nancy. He is member of the UK Serials Group, associate editor of TGJ, a member of the Editorial Board for Interlending & Document Supply, and a member of the former EAGLE, European Association for Grey Literature Exploitation.

Christiane Stock graduated from the University of Freiburg in 1984. She joined INIST-CNRS the French Institute of Scientific and Technical Information in 1989. Member of the Technical Committee for the SIGLE database since 1993, she also set up the national agency for the ISRN (International Standard Report Number). Today she is the head of the monographs and grey literature section at INIST.

Albert K. Boekhorst (1943) studied sociology at the University of Amsterdam. Since 1987 he is working at the dept of Information Science at the same university. He is visiting professor at the University of Pretoria (South Africa) and also teaches at the University of Peking (China) and at the Pedagogical University of Tallinn (Estonia). He teaches courses on the role and infrastructure of information and knowledge in organisations. In his research he concentrates on theoretical aspects of the role of information in societies in general and the access to information more specifically. Information literacy is a special field of interest in this context. More information on his work is to be found at www.hum.uva.nl/akb. Comments to boekhorst@uva.nl